Available online at www.sciencedirect.com

ScienceDirect

Control Engineering Practice 15 (2007) 1001–1011

CONTROL ENGINEERING PRACTICE

www.elsevier.com/locate/conengprac

Application of a distributed large basin runoff model in the Great Lakes basin

Chansheng He[a,b,*], Thomas E. Croley II[c]

[a]Department of Geography, Western Michigan University, Kalamazoo, MI 49008, USA
[b]Cold and Arid Regions Environmental and Engineering Research Institute, The Chinese Academy of Sciences, Lanzhou, Gansu 730000, PR China
[c]Great Lakes Environmental Research Laboratory, NOAA, Ann Arbor, Michigan 48105, USA

Received 16 May 2005; accepted 29 January 2007
Available online 19 March 2007

Abstract

This paper analyzes the application of a spatially distributed large basin runoff model (DLBRM) in the Great Lakes Basin of the United Stats and Canada and discusses four essential components of operational hydrologic model development: model structure, model input, model calibration, and Geographical Information System (GIS)-model interface. The results indicate that large scale operational hydrologic models that are based on mass continuity equations and include land surface, soil zones, and groundwater components require fewer parameters, are less data demanding, and are particularly suitable for solving water resources problems over large spatial and temporal scales than many other models. Use of GIS-model interfaces is essential for utilizing the existing multiple digital databases in defining model input and in facilitating model implementation and applicability.

Keywords: Hydrologic modeling; Great Lakes; Spatial variability; Calibration; GIS interface

1. Introduction

Simulation models are useful tools in hydrologic research (e.g. research hypothesis testing and understanding of hydrologic processes), water resources planning (e.g. floodplain assessment and ecosystem protection), and management (e.g. water resource allocation and soil erosion control) (Castelletti, de Rigo, Rizzoli, Soncini-Sessa, & Weber, 2006; Duviella, Chiron, Charbonnaud, & Hurand, 2007). With rapid advances in the availability of digital databases and computing technology, numerous hydrologic models have been developed during the past three decades. These models can be generally grouped subjectively into three types: empirical/statistical (input–output or black-box), physically based, and conceptual. Empirical/statistical models seek to represent hydrologic system response by extracting information from existing databases (an input–output description of the phenomenon) without considering any of the physical processes involved (Bras, 1990; Kokkonen & Jakeman, 2001; Merritt, Letcher, & Jakeman, 2003). Such models are simple, less data demanding, and are particularly useful for understanding hydrologic responses on a spatially lumped and/or temporally coarse resolution (Merritt et al., 2003).

Physically based models use the conservation of mass and momentum equations to describe hydrologic processes of a watershed, e.g. the model by Freeze (1972), the Système Hydrologique Europèen (SHE) model, and the THALES model (Beven, 2000). Such models exclusively describe all the important rainfall runoff components and processes by a set of differential equations. Solving those equations requires a huge amount of computing power and quite detailed databases for specifications of many parameter values over each of the elements in the solution domain (Beven, 2000; Merritt et al., 2003).

Conceptual models describe all of the component hydrologic processes as a system of interconnected storages that are recharged and depleted in accordance with mass continuity equations, without trying to be exact representations of

*Corresponding author. Department of Geography, Western Michigan University, Kalamazoo, Mi 49008, USA. Tel.: +1 269 3873425; fax: +1 269 3873442.
E-mail address: he@wmich.edu (C. He).

0967-0661/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.conengprac.2007.01.011

physical reality (Bras, 1990; Beven, 2000; Kokkonen & Jakeman, 2001; Croley & He, 2005). These models use a set of distribution functions (e.g., statistical, simple functional forms, unit hydrographs) to represent runoff generation processes, rather than full process descriptions, across the study watershed (Beven, 2000). These models are able to reflect the hypothesis about the processes governing system behavior (Merritt et al., 2003). Examples of such models include the Stanford Watershed Model (Crawford & Linsley, 1966), Mike-11 (Danish Hydraulic Institute, 1990), the Hydrologic Simulation Program in FORTRAN (Bicknell, Imhoff, Kittle, Donigian, & Johansen, 1996), and the large basin runoff model (LBRM) (Croley, 2002).

Traditionally, empirical/statistical and conceptual hydrologic models have treated input parameters as lumped over the entire study watershed and ignored spatial variability of hydrologic processes. While still appropriate for simulating hydrologic processes and outputs at the watershed outlet, lumped models ignore the spatial variability of watershed processes (Merritt et al., 2003). With fast development in digital databases and computing technology and increasing requirements for knowledge of distribution of water and material transport, distributed models (both physically based and conceptual) have been increasingly applied to understand the spatial and temporal variations of watershed processes and outputs (Merritt et al., 2003). Distributed, physically based models such as SHE are primarily designed to simulate hydrologic processes in great detail at the micro scale (usually $<10^2\,\mathrm{km}^2$) and require multiple massive amounts of data to compute spatial and temporal distributions of energy and water balances in the soil–plant–atmosphere system. Such detailed input data are often expensive and difficult to collect over large watersheds (say 10^3–$10^4\,\mathrm{km}^2$). As a large watershed may be discretized into thousands of grid cells, this type of model requires much computational power, challenging even with current computational technology (Bras, 1990; Beven, 2000; Kokkonen & Jakeman, 2001; Croley & He, 2005). Also, with many different parameters involved for each cell, parameter calibration becomes extremely difficult.

Facing the tradeoff between the representation of hydrologic processes and the availability of input data, distributed conceptual models have been developed to predict hydrologic phenomena (e.g. forecasting lake level fluctuations or estimating surface runoff) over a large geographic area (usually $>10^3$–$10^4\,\mathrm{km}^2$) at long time scales (typically for use over monthly, annual, or longer time scales at a daily interval) (Croley & He, 2005). Such models include the Areal, Nonpoint Source Watershed Environment Response Simulation (ANSWERS) (Beasley, Huggins, & Monke, 1980), the identification of hydrographs and components from rainfall, evaporation and streamflow data (IHACRES) (Jakeman & Hornberger, 1993), the US Geological Survey's Precipitation-Runoff Modeling System (PRMS) (Leavesley & Stannard, 1995), and the variable infiltration capacity (VIC) model (Liang, Lettenmaier, Wood, & Burgs, 1994). Implementations of such models are often constrained by limited data availability, computational requirements, and model application costs over larger areas (Croley & He, 2005). While it is essential for them to be able to simulate spatial distributions of hydrologic processes, large-scale hydrologic operational models (models that are used to predict hydrologic phenomena regularly and routinely over a large geographic area) must have few parameters, use easily accessible multiple digital databases, and be easy to use in order to support water resources applications (e.g. water supplies, water quality management, navigation, and irrigation) over large areas under a wide range of climates (Rango & Shalaby, 1998).

This paper describes application of a spatially distributed, two-dimensional, conceptual model that is based on the equations of conservation of mass for runoff generation for supporting hydrologic applications over large watersheds in the Laurentian Great Lakes Basin. It first reviews recent developments in hydrologic modeling and then discusses four essential components: model structure, model input, model calibration, and geographical information system (GIS)-model interface, in the development and implementation of large scale hydrologic models. Finally, it describes the application of those components in the application of the two-dimensional distributed large basin runoff model (DLBRM) to the Great Lakes watersheds to demonstrate incorporation of these issues in the model development process.

2. Recent developments in operational hydrologic modeling

Significant progress has been made in hydrologic modeling during the past three decades. The following sections give a brief review on recent developments in model input, model structure, model calibration, and GIS-model interface for large operational hydrologic modeling.

2.1. Utilization of multiple GIS and remote sensing databases

Rapid advances in remote sensing, GIS, digital databases, and computing technology during the past three decades have provided enormous opportunities for the hydrologic research community. For example, newly launched satellites, such as the Earth Observing System (EOS) PM-1, RADARSAT (space borne radar), LANDSAT 7 Enhanced Thematic Mapper (TM) Plus, Space Imaging, Inc's 1 m resolution of the IKONOS satellite, and others, enable the extraction of hydrologic parameters (e.g. areal estimates of precipitation, snow water equivalent and snow cover extent, vegetative cover, surface temperature, surface albedo, and incoming solar radiation, soil moisture, etc) over multiple temporal and spatial scales. Digital Elevation Model (DEM) databases are widely used for deriving slope, aspect, drainage network, and flow direction for a watershed (Hornberger & Boyer, 1995). Soil databases such as the State Soil Geographic Data Base (STATSGO) from the US Department of Agriculture Natural Resource Conservation Service (NRCS) (1994) enable the incorporation of spatial variation of soil characteristics into hydrologic models (He, Riggs, & Kang, 1993; He, Shi, Yang, & Agosti, 2001). Land cover databases allow the derivation of land use/cover related parameters such as leaf area index, zero plane displacement height and Manning's coefficient values to hydrologic models.

While availability of a large number of digital databases makes extraction of some model input variables (e.g. land cover) possible over large areas, obtaining certain input variables for operational hydrologic models, especially for spatially distributed models, remains a challenge. For example, precipitation is a key parameter in rainfall-runoff modeling. Estimates of the spatial distribution of precipitation are still inadequate due to a lack of spatial and temporal coverage of satellites and rain gauge stations, particularly in rural areas. Methods for estimating precipitation rates, such as cloud indexing, thresh-holding, and life history methods, by satellite remote sensing are still at an experimental stage (Engman, 1995). Microwave and geosynchronous orbiting satellites such as Geostationary Operational Environmental Satellite (GOES) can only provide limited types of observations (Engman & Gurney, 1991; Engman, 1995). Ground-based radar is currently limited to a measurement circle with a radius up to about 100 km and its distribution is mainly limited to densely populated areas (Engman & Gurney, 1991; Engman, 1995). Estimates of precipitation from those radar stations still need to be calibrated against measurements from nearby rain gauges. Application of satellite remote sensing is still at a research stage for estimating soil moisture and determining sediment load. No remote sensing methods have been found to measure streamflow in river basins, or infiltration of precipitation into the soil, deep soil moisture or groundwater, or the levels of chemical pollutants in water bodies (Rango, 1994). Thus, operational hydrologic models for large basins must still rely on inadequately distributed rain gauges for estimates of precipitation. In addition, unlike precipitation networks, there are virtually no systematic measurements of solar radiation and surface temperature throughout the US Although algorithms are available to derive solar radiation and surface temperature from visible and thermal bands of satellites such as GOES, LANDSAT TM, and Advanced Very High Resolution Radiometer (AVHRR) (Hall, Huemmerich, Goetz, Sellers, & Nickeson, 1992; Lindsey & Farnsworth, 1997), application of those algorithms often requires knowledge and skills of image processing and interpretation. Therefore, development of large-scale operational hydrologic models need to take advantage of opportunities provided by remote sensing and GIS databases, and at the same time, to consider limitations of data availability in mathematical formulation and parameter specifications.

2.2. Structure of operational models

Large-scale operational models represent all the component hydrologic processes as a system of interconnected storages with mass continuity equations. Model components should include land surface, soil zones, and groundwater to produce realistic estimates of rainfall-***runoff generation (Koren, Finnerty, Schaake, Smith, Seo, Duan, et al., 1999; Martinez, Dunchon, & Crosson, 2001). Variable-source-area concepts (runoff from a dynamically changing surface area) should be used in computing infiltration and saturation runoff as the variable-source models give a better representation of hydrologic processes and produce better estimates of overland flow and are less scale dependent (Quinn, Beven, & Culf, 1995; Abdulla, Lettenmaier, Wood, & Smith, 1996; Koren et al., 1999; Beven, 2000; Valeo & Moin, 2001). Soil layers and groundwater should be included in the model structure as water budget is very sensitive to the number of layers in the soil profile and omission of the subsurface-groundwater component in a runoff model can lead to an increase in model scale dependency (Koren et al., 1999; Martinez et al., 2001). For modeling soil water storage, a single layer in both the upper and bottom soil zones is adequate (Martinez et al., 2001).

The Penman-Monteith (PM) method and the complementary relationship (CR) methods have both been widely used for estimating regional evapotranspiration (ET) over long periods of time. The PM method assumes that actual ET does not affect potential ET (the ET and potential ET are "independent"). It links the effects of vegetation to the ET process through aerodynamic and canopy resistance terms and may be more appropriate for small areas where detailed databases are available. The CR concept states that as water availability becomes limited then actual ET falls below its potential, and an excess amount of energy becomes available in the form of sensible heat and/or long-wave back radiation that increases the temperature and humidity gradients of the over passing air and leads to an increase in potential ET equal in magnitude to the decrease in ET. If water availability is increased, the reverse occurs, and ET increases as potential ET decreases. Thus, potential ET can no longer be regarded as an independent causal factor. Instead it is predicated upon the prevailing conditions of moisture availability (Hobbins, Ramirez, Brown, & Claessens, 2001). The CR methods bypass complex and poorly understood soil–plant interactions, require fewer parameters for applications, and may be more applicable to large areas where detailed data sets are not available (Hobbins et al., 2001; Liang et al., 1994; Morton, 1994; Silberstein, Sivapalan, & Wyllie, 1999; Sugita, Usui, Tamagawa, & Kaihotsu, 2001; Xu & Singh, 1998). Dependent upon the data availability and the modeling scale, either method may be used in large-scale operational models.

Spatial variations of climate and landscape have significant impacts on runoff modeling (Beven, 2000).

Operational models should take advantage of available databases of DEM, vegetation, soil, climate, and hydrography to account for spatial variations of climate, soil, topography, and land use practices. Dependent upon the purpose of the study and data availability, the study watershed should be discretized into either grid network or hydrologic response units (HRUs), large-scale operational models applied to each cell or HRU, and output from each cell routed to the watershed outlet for identifying and understanding of both hydrologic responses and their spatial distribution in the study watershed (Becker & Braun, 1999; Karvonen, Koivusalo, Jauhiainen, Palko, & Weppling, 1999).

2.3. Model calibration and uncertainty

Hydrologic models must be calibrated (model parameters estimated) to well represent reality, i.e. to match observations with acceptable accuracy and precision (Gupta, Sorooshian, & Yapo, 1998; Loagues & Kyriakisdis, 1997). Traditionally, hydrologic model calibration is done with split-sample testing by using streamflow data to find the "best" parameter set (Mroczkowski, Raper, & Kuczera, 1997; Gupta et al., 1998). This approach is practical, less data demanding, and easier to implement than dealing with multiple-response data. Recently researchers have explored different calibration approaches. Mroczkowski et al. (1997) state that use of multiple-response data (e.g. streamflow, soil moisture, and chemical stream loads) in a watershed experiencing a change in hydrologic regime gives a better assessment of the model structure than the traditional split-sample testing using streamflow data alone in undisturbed watersheds. However, obtaining independent, multiple-response data may be very difficult, particularly over large watersheds. Others (Yan & Haan, 1991; Loagues & Kyriakisdis, 1997) suggest use of a multiobjective approach in model calibration for better assessment of the limitations of model structure and confidence of model predictions when multiple objectives cannot be easily transformed into a single common objective due to a lack of quantitative comparison measures. But nonlinearity in parameters and multiple optima make calibration by multiple criteria very difficult. A residual-based approach is likely to have similar power but is easier to implement (Mroczkowski et al., 1997).

The inherited uncertainties in model input data and parameters affect model performance even after calibration. Inadequate consideration of the spatial variability of precipitation data introduces greater uncertainty into parameter estimates than errors in runoff data (Borah & Haan, 1991; Chaubey, Haan, Grunwald, & Salisbury, 1999) and inaccuracies in DEM and the DEM-derived drainage network affect estimates of runoff peaks, timing and volume (Holmes, Chadwick, & Kyriakidis, 2000; and Kenward, Lettenmaier, Wood, & Fielding, 2000). Merritt et al. (2003) state that complex models (physically based and conceptual models) are more vulnerable to problems of parameter identifiability (behavior of the model not observed in a particular sample of data or multiple behavior types of the model cannot be differentiated by the observed data). Gan, Dlamini, & Biftu (1997) evaluate five conceptual rainfall-runoff models of different complexity (ranging from 9 to 21 parameters) and report that model performance is more associated with the model structure, the objective function in calibration, and data quality and less related to model complexity or calibration data length. Some have investigated uncertainties of hydrologic models by using methods such as Monte Carlo based approaches and the mean-value first-order second-moment method (Chaubey et al., 1999; Beven, 2000; Muleta & Nicklow, 2005). Others state that performance of the spatially distributed models can only be assessed with spatially distributed observations that are technically not feasible, especially over large areas (Merritt et al., 2003). Considering the limitations of data, the criteria for model acceptability need to be carefully defined, since if the criteria are too strict, all models will be rejected (Beven, 2000).

2.4. GIS-model interface

Development of operational hydrologic models, particularly distributed hydrologic models, requires integration of GIS, remote sensing, and other digital data bases for extracting the needed model variables, and for processing, analyzing, and visualizing the model results (He, 2003). A number of GIS-model interfaces have been developed to assist users in data organization, parameter extraction, model execution, and output display, and to improve model applicability. Such interfaces include linkages between Geographic Resource Analysis and Support System (GRASS) and AGNPS (Agricultural Nonpoint Pollution Model) (He et al., 1993), and Arc/Info and HEC-HMS (Hydrologic Modeling System) (Hellweger & Maidment, 1999). He, Shi, Yang, & Agosti (2001) developed an interface to integrate the ArcView GIS and AGNPS for modeling and analysis of agricultural watersheds. A software package, Real-time Interactive Basin Simulator (RIBS) by Garrote & Bras (1995) integrates a radar-based rainfall prediction model, a DEM-based rainfall-runoff model, and other multiple databases to forecast real-time flooding. Since the 1990s, the US Environmental Protection Agency (2001) has developed and updated Better Assessment Science Integrating Point and Nonpoint Sources (BASINS) system to incorporate ArcView GIS and hydrologic models in support of water quality programs nationwide. To better represent hydrologic processes and facilitate model implementation and applicability, operational hydrologic models should incorporate linkages or interfaces to GIS for data integration, analysis and visualization.

3. Analysis of spatially distributed large basin runoff model

The Laurentian Great Lakes are the largest surface freshwater system in the world and are vital to the economic prosperity of both the US and Canada. This interconnected system supports multiple water uses that include navigation, hydropower generation, agricultural production, urban development, tourism and recreation, and fishery and wildlife habitat in the region (He, 1997). To make sustainable use of this precious fresh water system, a large-scale operational hydrologic model, the Large Basin Runoff Model (LBRM) has been developed since the 1980s to support water resource decision-making (Croley, 2002). However, the LBRM is a lumped (1-D), conceptual parameter model. Even though able to simulate integrated hydrologic responses at the watershed outlet and having been successfully applied to each of the 121 watersheds in the Great Lakes Basin for multiple water resources applications since the 1980s, it does not take into account spatial heterogeneity and thus is unable to simulate the spatial response to hydrologic events. To overcome this limitation, the lumped LBRM is modified into a spatially distributed LBRM (DLBRM) to help researchers and resource planners to better understand the spatial distribution of hydrologic processes and their response. The following sections briefly describe the structure, input, calibration, and GIS interface of the DLBRM.

3.1. Lumped LBRM

The LBRM (lumped 1-D) was developed by the National Oceanic and Atmospheric Administration (NOAA)'s Great Lakes Environmental Research Laboratory (GLERL) in the 1980s to support hydrologic simulations and water resources applications in the Great Lakes Basin. It uses a serial and parallel cascade of linear reservoirs (outflows proportional to storage) to represent moisture storages within a watershed: surface, upper soil zone (USZ), lower soil zone (LSZ), and groundwater zone (GZ) (Croley & He, 2005). The model computes the total heat available each day, indexed by daily air temperature, to become potential evapotranspiration (ETP) or actual evapotranspiration (ET), a complementary approach. It splits the heat available between ETP and ET, by preserving the total heat and taking ET as proportional to both ETP and storage. The model uses variable-area infiltration (infiltration proportional to the unsaturated fraction of USZ) and daily precipitation and degree-day snowmelt (Croley, 2002). It has been applied extensively to the 121 riverine watersheds draining into the Laurentian Great Lakes for use in both simulation and forecasting (Croley, Quinn, Kunkel, & Changnon, 1998; Croley & He, 2005). The LBRM uses readily available climatological data, requires few parameters and data, uses mass continuity equations to govern the water storages in each storage zone, and is generally applicable to large watersheds (Croley, 2002). However, it does not consider spatial distribution of hydrologic responses and thus limits its applications in water resources operations and management.

Recently, the LBRM was modified from its aggregated-parameter definition for an entire watershed to a two dimensional representation of the 1-km^2 flow cells comprising the watershed (Croley & He, 2005, 2006). The structure, input, calibration, and GIS-interface of the DLBRM are briefly described below.

3.2. Model structure

The structure of the DLBRM is shown in Fig. 1. The continuity equations of the LBRM were modified to allow upstream surface and subsurface routing between cells of flows of the surface zone, the USZ, the LSZ, and the GZ. This enables surface and subsurface storages to interact both with each other and with adjacent-cell surface and subsurface storages. The watershed is discretized into a grid network of 1 km^2 cells. Since ET and potential ET cannot be regarded as complementary when the DLBRM is applied to a small cell, both are replaced with the more traditional "independent" concept (that actual ET does not affect potential ET) in the DLBRM. A flow network is generated by identifying the network flow cascade of the watershed cells based on the slope, flow directions, and receiving cell numbers (the cell that receives the upper stream flow) derived by the GIS interface. A flow hydrograph out of a cell is saved as a "pending" inflow hydrograph into the next downstream cell, until all upstream inflows for that next cell are computed; then these flows are added together to determine the total upstream surface flow into that next cell [details of the DLBRM are described by Croley and He (2005, 2006)].

3.3. Model input and output

Input variables to the DLBRM include, for every cell in the watershed grid, daily precipitation and air temperature, solar isolation, elevation, slope, flow direction, land use, depths (cm) of USZ and LSZ, available water capacity (%) of USZ and LSZ, soil texture, permeability (cm/h) of USZ and LSZ, Manning's coefficient values, and daily flows (see Tables 1 and 2). Precipitation and temperature are interpolated from more than 1800 historical climatological site records in the Great Lakes Basin by using one of several methods including: Thiessen polygon, inverse distance, inverse squared-distance, and linear interpolation over a triangular irregular network (TIN). Daily surface insolation estimates are generated by two methods: (1) from temperature databases by empirical formulae, and (2) reversed-engineered from an available weather generation model as a function of location, day of the year, air temperature and precipitation (Croley & He, 2005). Slope and flow direction are extracted from digital elevation model (DEM) databases. USZ and LSZ depths, available water capacity, permeability, and soil texture are derived

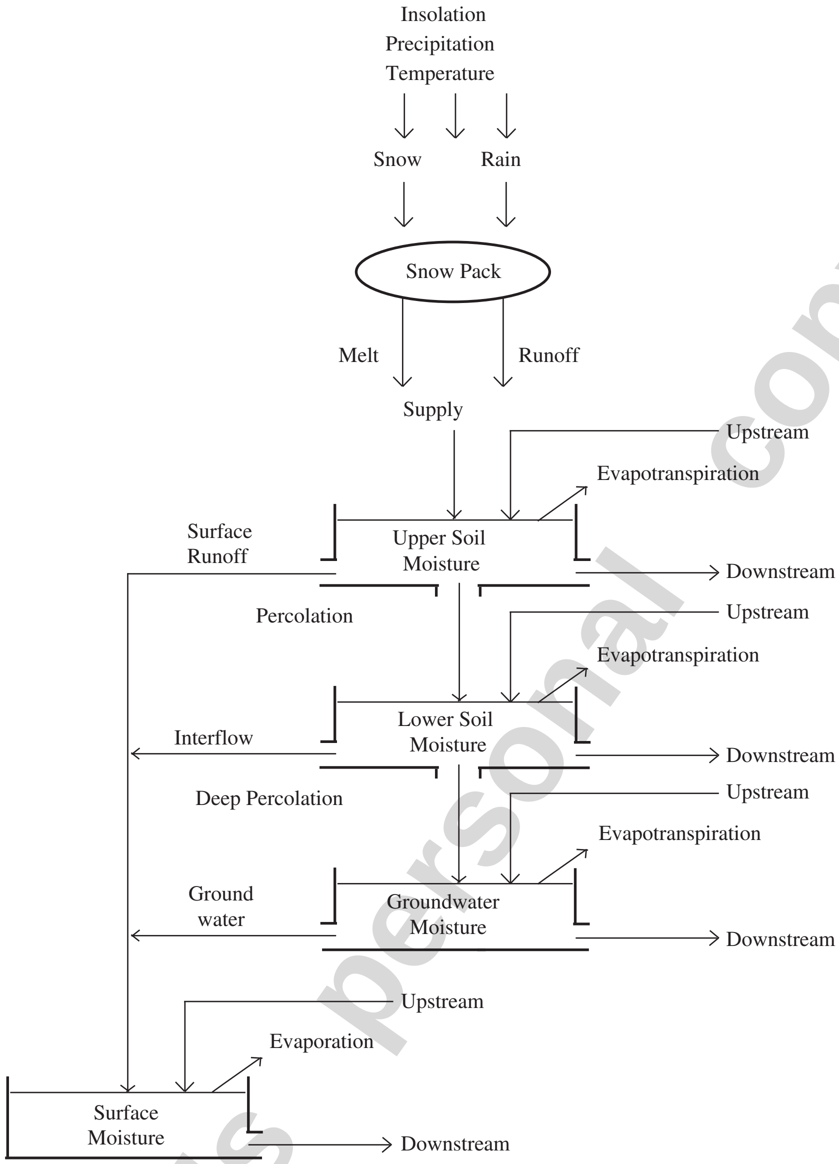

Fig. 1. Tank cascade schematic of distributed large basin runoff model.

from STATSGO (US Department of Agriculture, 1994). Manning's coefficient values are derived for each grid based on the combination of land use, slope, and soil texture.

DLBRM output includes, for every cell in the watershed grid, surface runoff to surface storage, infiltration to USZ, ET, ETP, percolation from USZ to LSZ, interflow from LSZ to surface storage, deep percolation from LSZ to groundwater storage, groundwater flow from groundwater storage to surface storage, surface moisture storage, USZ, and LSZ moisture storages, groundwater storage, and lateral flows from storages to adjacent cells for the surface (channel outflow), USZ, LSZ, and groundwater. Currently, daily precipitation and air temperature and solar insolation to the DLBRM still rely on measurements from ground-based weather stations. Once the daily, areal coverage of snow pack, rainfall, and solar radiation (from remote sensing sensors such as NOAA, GOES, and other EOS satellites) become available on a routine basis, the DLBRM can be modified to utilize these estimates to simulate rainfall-runoff for the Great Lakes Basin. Such addition will lead to better representation of the spatial distribution of net supply to the model and hence significantly improve the accuracy of the runoff simulation.

3.4. Model calibration

Availability of detailed observation data may be more important than more complex calibration methods. Muleta and Nicklow (2005) applied a hierarchy of three techniques (screening, parameterization, and sensitivity analysis) in calibrating soil and water assessment tool (SWAT) (35 parameters) but the simulated streamflow and sediment concentrations showed large discrepancies compared to

Table 1
Input variables for the DLBRM

Variables	Databases
Elevation	USGS digital elevation model (DEM)[a]
Flow direction	USGS DEM
Slope	USGS DEM
Land use	USGS land use database[b]
Depth of upper soil zone (USZ)	USDA STATSGO[c]
Depth of lower soil zone (LSZ)	USDA STATSGO
Available water capacity (%) of USZ	USDA STATSGO
Available water capacity of LSZ	USDA STATSGO
Permeability of USZ	USDA STATSGO
Permeability of LSZ	USDA STATSGO
Soil texture	USDA STATSGO
Manning's coefficient value	Land use, slope, and soil texture

Sources:
[a]US Geological Survey National Elevation Dataset (NED) http://seamless.usgs.gov/.
[b]US Geological Survey National Landcover Characterization Dataset (NLCD) 1992, http://seamless.usgs.gov/.
[c]US Department of Agriculture 1994. http://soils.usda.gov.

Table 2
Time series meteorological and flow variables for the DLBRM

Variables	Databases
Daily precipitation	National weather service climate databases
Daily air temperature	National weather service climate databases
Daily solar isolation	National weather service climate atlas
Daily flows	USGS discharge database

observations at the outlet. They attribute the discrepancies to the lack of detailed observations in the study watersheds. Considering the availability of hydrologic data, computation time, and long term simulation period (multiple decades) over large-scale watersheds, calibration of the DLBRM is conducted as a systematic search of the parameter space to minimize the root mean square errors (RMSE) between actual and simulated daily outflow volumes at the watershed outlet (note: two other indices: correlation coefficient and the coefficient of efficiency or Nash–Sutcliffe coefficient are also available for use in the DLBRM automatic calibration program).

There are 15 parameters in the model: degree-day snowmelt coefficient, heat-temperature index coefficient, USZ capacity, partial linear reservoir coefficients for surface zone evaporation, USZ ET, LSZ ET, and GZ ET, linear reservoir coefficients for percolation, interflow, deep percolation, groundwater flow, and lateral downstream surface flows, lateral downstream USZ flows, lateral downstream LSZ flows, and lateral downstream groundwater flows. While parameters describing the degree-day snowmelt and heat-temperature index were taken as spatially constant, and while surface and groundwater evaporation were taken as zero, the spatial structures of other parameters were assigned as follows. The coefficients for percolation, lateral downstream USZ flow, and USZ ET were taken proportional to observed USZ permeability; the USZ capacity was taken proportional to observed USZ available water capacity; the coefficients for interflow, deep percolation, lateral downstream LSZ flow, LSZ ET, groundwater flow, and lateral downstream groundwater flow, were taken proportional to observed LSZ permeability; the coefficient for lateral downstream surface flow was taken proportional to the square root of surface slope divided by Manning's coefficient.

While spatial variation in each of the parameters was described as above, the spatial average of each parameter was determined by minimizing the RMSE between model and observed watershed outflow through a systematic search of the 15-parameter space.

3.5. GIS interface

DLBRM use requires the watershed be discretized into a grid network at a resolution of 1 km^2 (to match existing areal coverage of meteorological data) or other sizes as specified by a user. It requires 15 input variables for each of the grid cells. To facilitate the implementation of the DLBRM, a GIS interface, the AVDLBRM (ArcView Distributed Large Basin Runoff Model) interface based on the work of He et al. (2001), has been developed for processing, extracting, analyzing, and visualizing model input and output. The interface consists of six modules: a soil processor, DLBRM utility, parameter generator, output visualizer, statistical analyzer, and land use simulator.

The soil database processor automatically derives, from STATSGO, spatially averaged depth, available water capacity (AWC), soil texture, soil slope, and permeability for the USZ (layer 1 in STATSGO) and LSZ (layers 2–6 in STATSGO) by soil association (soil association is a unit on which soil information is mapped and assembled). The parameter generator module helps a user first set up input files of DEM, land use, soil, and hydrography, and then derives required input parameters of slope, receiving cell numbers, flow direction, depths of USZ and LSZ, AWC of USZ and LSZ, permeability of USZ and LSZ, soil texture, and elevation for each grid cell. As the flow net allows only one outlet for the entire study watershed, flow directions must be carefully inspected to eliminate any flow loops. A DLBRM utility module in the AVDLBRM is used to check such errors and allows the user to edit flow direction either one cell at a time or by several cells at a time (He, 2003). The verified flow net is then used to route flow (Croley & He, 2005).

To derive appropriate Manning's coefficient values for each grid cell, the interface first helps the user define the hydrologic response units (HRUs) based on combinations of land use, slope, and soil texture (i.e. dividing the watershed into different HRUs) for the entire watershed and then uses a look-up table to assign each HRU an appropriate Manning's coefficient value automatically. These values are mainly determined by land use/cover categories and then adjusted by slope and soil texture. Subsequently, the interface assigns values from each of HRU to each grid cell.

The output visualizer allows the user to select any variable from the output file and display it in map format in ArcView. A separate animation program has also been developed to animate the output variables for multiple years at daily intervals, which enables examination of dynamics of hydrologic variables over the long term. The statistical analyzer enables the user to conduct an analysis of variance (ANOVA) to examine relations between land use/cover and simulated results. The land use change simulator allows the user to specify land use change scenarios in a sub-basin or specific area based on the land use/cover file and evaluate the hydrologic impact of such changes to the downstream area (He, 2003).

Compared to the lumped parameter LBRM, the DLBRM has the following improvements: (1) two-dimensional representation of the 1-km^2 flow cells comprising the watershed; (2) consideration of spatial variation of climate, soil, topography, and land use in simulating watershed hydrologic processes; (3) incorporation of impacts of upper stream landscape heterogeneity to downstream hydrologic responses by routing flows from USZ, LSZ, GZ, and surface storages downstream; and (4) better representation of the watershed hydrologic system over the large scale. Moreover, the continuity equations in DLBRM are solved analytically, thus minimizing computational errors and enabling the tracking of hydrologic processes. Furthermore, the representation of the watershed system by parallel cascades of linear reservoirs makes the DLBRM less data demanding and suitable for large watershed applications.

4. A DLBRM application example

The DLBRM has been applied to 18 watersheds with drainage areas ranging from 3000 to over 24,000 km^2 in the Great Lakes Basin. This section describes its application to the Kalamazoo River Watershed (drainage area of approximately 5612 km^2) in southwest Michigan of the US. The watershed was discretized into a grid network of 5612 cells at 1 km^2 resolution. A DEM from the USGS (http://seamless.usgs.gov/) was used to derive topographically related parameters (flow direction, receiving cell number, and slope) by the AVDLBRM interface. The land cover data from the 1992 USGS National Land cover Characterization Data set (NLCD) (http://seamless.usgs.gov/) was used to derive a land cover category (code) for each grid cell. The STATSGO dataset was used to derive depth (cm), AWC (cm), soil texture, and permeability (cm/h) for the USZ and LSZ for each of the cells. Manning's coefficients were assigned to each cell by the HRUs by the AVDLRBM interface. Average daily river flow rates from the USGS were converted into daily outflow volumes and were used to conduct a systematic search of the parameter space to minimize the RMSE between actual and simulated daily outflow volumes at the watershed outlet (Croley, He, & Lee, 2005).

The DLBRM was calibrated over the period of 1950–1964 for each of the 5612 cells (1-km^2) at a daily interval. The calibration took over a week on a desktop personal computer (Intel Xeon^{TM} processor @ 2.40 GHz). The calibration shows a 0.88 correlation between simulated and observed watershed outflows (0.77 coefficient of determination), a 0.020 cm/d RMSE (0.57 RMSE divided by standard deviation), and a coefficient of variation of 0.24. Over a separate verification period (1965–1990), the model demonstrated a 0.81 correlation between simulated and observed watershed outflows (0.66 coefficient of determination), a 0.028 cm/d RMSE (0.58 RMSE divided by standard deviation), and a coefficient of variation of 0.50. Comparison of the lumped parameter LBRM and distributed parameter DLBRM calibration statistics for the period of 1950–1964 shows that both models show very similar ratios of model to actual flow means and very similar long-term average ratios to surface supply of surface runoff, interflow, groundwater flow, USZ ET, and LSZ ET even though the DLBRM has a slightly higher RMSE than the lumped LBRM. The slightly higher RMSE in DLRBM is expected since spatial landscape heterogeneity of the watershed (Table 3) and spatial meteorology over the watershed is estimated more poorly than spatially average values. Spatial landscape data exist at multiple spatial resolutions. Soil related input variables are derived from STATSGO, assuming uniform distribution of the variables within each soil association. Topographic variables such as slope and flow direction are extracted from the DEM (1:250,000 scale) and then aggregated to the 1 km^2 cell by selecting dominant values. This aggregation process inevitably generates errors into the hydrologic models like DLBRM (Holmes et al., 2000 and Kenward et al., 2000). Daily precipitation and temperature input to

Table 3
Comparison of selected summary statistics for lumped-parameter LBRM and distributed LBRM calibration in the kalamazoo river watershed

Calibration period 1950–1964	Correlation	RMSE (cm)	Long term average ratio to surface supply					
			Model outflow	Surface runoff	Interflow	Groundwater flow	Upper zone ET	Lower zone ET
Lumped	0.895	0.0184	1.005	0.068	0.032	0.246	0.574	0.074
Distributed	0.880	0.0217	1.004	0.061	0.072	0.222	0.578	0.072

the DLBRM are spatially interpolated (using Thiessen polygon, inverse distance, and inverse squared-distance methods) from a network of over 1800 weather stations in the Great Lakes Basin, corresponding to about 290 km^2/station. Such lack of detailed spatial representation of meteorological data leads to large uncertainties in parameter estimates and hence model output (Chaubey et al., 1999; Vachaud & Chen, 2002).

Comparison of hydrographs for the period of 1950–1952 shows overall good agreement (Fig. 2). The water balance shows an absence of storage in the lower soil zone, implying that the groundwater zone receives its input directly from the upper soil zone. The base flow seems well represented but several peak flows are under-estimated. Fig. 3 also shows surface runoff flows out of the USZ into the surface zone, groundwater flows from the GZ to the

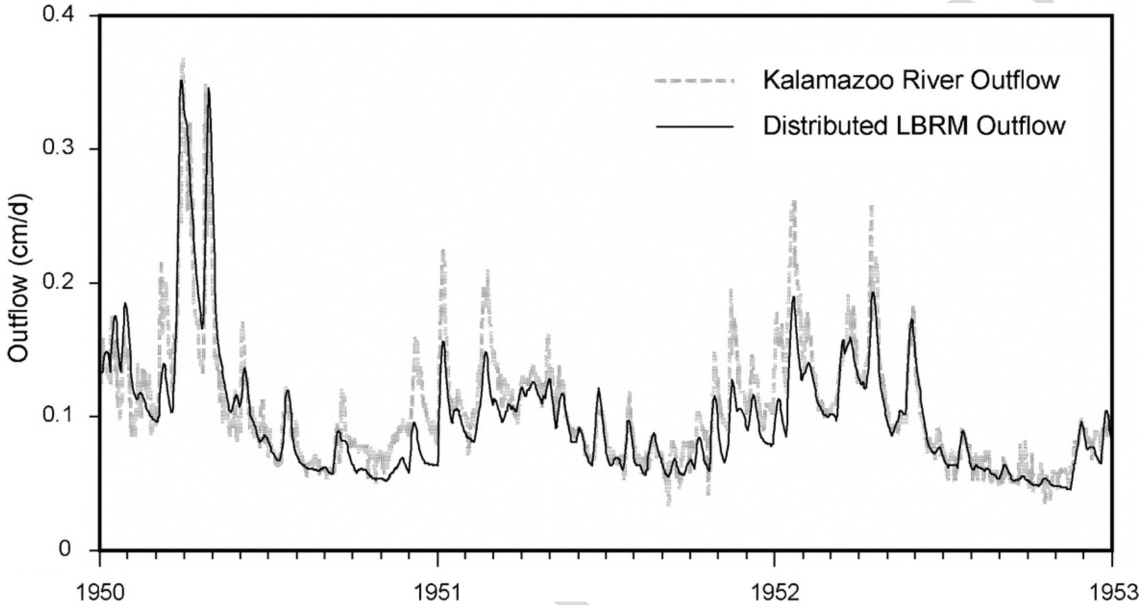

Fig. 2. Comparison of simulated river outflows with actual Kalamazoo River outflows for the period of 1950–1952.

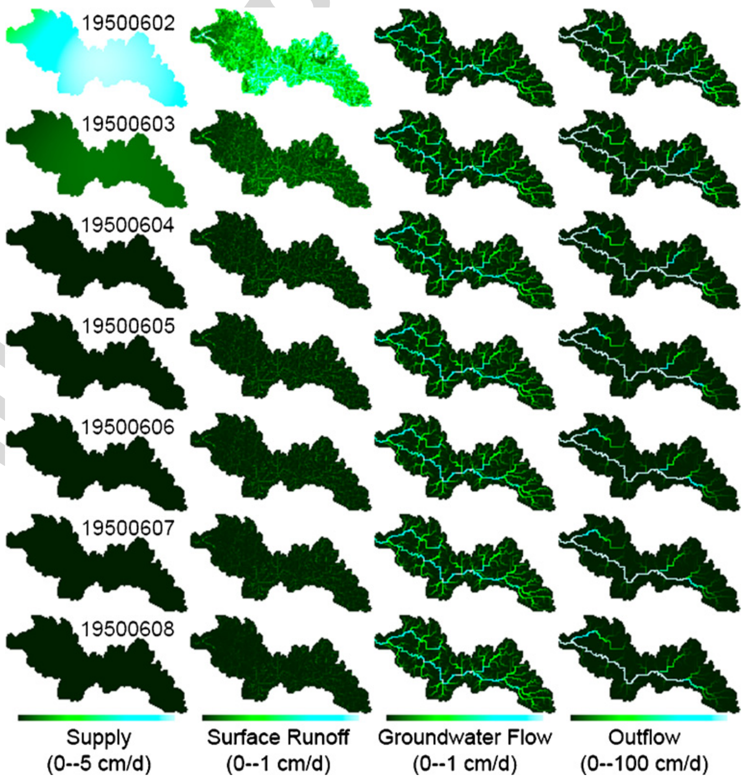

Fig. 3. Distributed large basin runoff model output for the Kalamazoo river watershed for the period of June 2–8, 1950.

surface, and outflow flows from the surface zone for the period of June 2–8, 1950. These flows represent the moisture storages within the watershed. The first flow is a within-the-cell flow while the last two cross cell boundaries and are accumulated down the flow network, reaching much larger values than within-the-cell flows. The general behavior of the watershed is depicted in this example. The supply on the first day results in a very flashy response in the upper soil zone, which is reflected by the immediate response in surface runoff. The groundwater zone is little affected and the groundwater flow is nearly constant throughout the period. The surface response lies in between; the outflow network is more dense at the beginning than at the end as water flows through the network throughout the period (Croley & He, 2005). The results of the DLRBM simulations show that the Kalamazoo River response to precipitation is dominated by groundwater storage, allowing delayed and sustained hydrologic responses to rainfall. These results are characteristic of the study watershed and are in agreement with other hydrologic studies of the Kalamazoo River such as Allen, Miller, and Wood (1972). The under-estimated peak flows probably result from poor estimation of "observed." The most downstream gauge station used in the calibration (USGS 04108500) has a drainage area of 4142 km^2 extrapolated to 5612 km^2 at the watershed outlet, which tends to impart a flashier response than actually exists at the watershed outlet.

These comparisons indicate that the DLBRM has indeed improved watershed representation and produced realistic simulations of hydrologic responses in the study watershed (Croley & He, 2005, 2006).

5. Summary

Through application of the DLBRM in the Great Lakes Basin, this paper discusses issues related to model structure, input, calibration, and GIS interface in large-scale hydrologic modeling. Both calibration and verification statistics for the DLBRM application to the Kalamazoo River watershed show good agreement between the model and observations. The results indicate that conceptual hydrologic models such as the DLBRM that are based on mass continuity equations and include surface, soil zones, and groundwater components, use fewer parameters, demand less data, require less computations, and better simulate large-scale hydrologic systems over the long-term than other models. Use of GIS is crucial in deriving input variables for each of the grid cells from the existing, readily available multiple digital databases and for facilitating and improving model implementation and application. While routing simulated flow downstream, considering surface and subsurface flow interactions between adjacent cells is essential for more accurate representation of the hydrologic processes. Calibration of the distributed hydrologic models at the watershed outlet, while less desirable, is reasonable and practical due to lack of detailed spatially distributed observations across the study watershed.

Acknowledgement

This is GLERL contribution number 1413. Partial support by Western Michigan University's Faculty Research and Creativity Fund and the Department of Geography Lucia Harrison Endowment Fund is appreciated.

References

Abdulla, F. A., Lettenmaier, D. P., Wood, E. F., & Smith, J. A. (1996). Application of a macroscale hydrologic model to estimate the water balance of the Arkansas-Red River Basin. *Journal of Geophysical Research*, *101*(D3), 7449–7459.

Allen, W. B., Miller, J. B., & Wood, W. W. (1972). *Availability of water in Kalamazoo County, Southwestern Michigan. US Geological Survey Water-Supply Paper 1973*. Washington, DC: US Government Printing Office.

Beasley, D. B., Huggins, L. F., & Monke, E. J. (1980). ANSWERS: A model for watershed planning. *Transactions of the ASAE*, *23*(4), 938–944.

Becker, A., & Braun, P. (1999). Disaggregation, aggregation and spatial scaling in hydrologic modeling. *Journal of Hydrology*, *217*, 239–252.

Beven, K. J. (2000). *Rainfall-runoff modelling: The primer*. New York: Wiley & Sons, Ltd. 360p.

Bicknell, B. R., Imhoff, J. C., Kittle, J., Donigian, A. S., & Johansen, R. C. (1996). *Hydrological Simulation Program—FORTRAN, User's Manual for Release 11*. Athens, Georgia: US Environmental Protection Agency, Environmental Research Laboratory.

Borah, A., & Haan, C. T. (1991). Impact of random data errors on parameter estimation in hydrologic modeling. *Transactions of the ASAE*, *34*(3), 857–864.

Bras, R. L. (1990). *Hydrology: an introduction to hydrologic science*. Reading, Massachusetts: Addison-Wesley Publishing Company 643p.

Castelletti, A., de Rigo, D., Rizzoli, A. E., Soncini-Sessa, R., Weber, E. (2006). Neuro-dynamic programming for designing water reservoir network management policies. *Control Engineering Practice*. Corrected Proof, Available online: 〈http://www.elsevier.com/locate/conengprac〉. doi:10.1016/j.conengprac.2006.02.011.

Chaubey, I., Haan, C. T., Grunwald, S., & Salisbury, J. M. (1999). Uncertainty in the model parameters due to spatial variability of rainfall. *Journal of Hydrology*, *220*, 48–61.

Crawford, N. H., & Linsley, R. K. (1966). *Digital simulation in hydrology: Stanford watershed model IV. Technical Report 39*. California: Department of Civil Engineering, Stanford University.

Croley II, T. E. (2002). Large basin runoff model. *Mathematical Models in Watershed Hydrology*, Water Resources Publications, Littleton, Colorado, pp. 717–770.

Croley, T. E., II, & He, C. (2005). Distributed-parameter large basin runoff model I: Model development. *Journal of Hydrologic Engineering*, *10*(3), 173–181.

Croley, T. E., II, & He, C. (2006). Watershed surface and subsurface spatial intraflows. *Journal of Hydrologic Engineering*, *11*(1), 12–20.

Croley, T. E., II, Quinn, F. H., Kunkel, K. E., & Changnon, S. J. (1998). Great Lakes hydrology under transposed climates. *Climatic Change*, *38*, 405–433.

Croley, T. E., II, He, C., & Lee, D. H. (2005). Distributed-parameter large basin runoff model II: application. *Journal of Hydrologic Engineering*, *10*(3), 182–191.

Danish Hydraulic Institute. (1990). *MIKE11 – A microcomputer based modeling system for rivers and channels: users manual*. Horsholm, Demark.

Duviella, E., Chiron, P., Charbonnaud, P., & Hurand, P. (2007). Supervision and hybrid control accommodation for water asset management. *Control Engineering Practice*, *15*(1), 17–27.

Engman, E. T. (1995). Recent advances in remote sensing in hydrology. *Reviews of Geophysics*, (Supplement), 967–975.

Engman, E. T., & Gurney, R. J. (1991). *Remote sensing in hydrology*. New York: Chapman and Hall.

Freeze, R. A. (1972). Role of subsurface flow in generating surface runoff. 2. Upstream source areas. *Water Resources Research*, *8*(5), 1272–1283.

Gan, T. Y., Dlamini, E. M., & Biftu, G. F. (1997). Effects of model complexity and structure, data quality, and objective functions on hydrologic modeling. *Journal of Hydrology*, *192*, 81–103.

Garrote, L., & Bras, R. L. (1995). An integrated software environment for real-time use of a distributed hydrologic model. *Journal of Hydrology*, *167*, 307–326.

Gupta, H. V., Sorooshian, S., & Yapo, P. O. (1998). Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information. *Water Resources Research*, *34*(4), 751–763.

Hall, F. G., Huemmerich, K. F., Goetz, S. J., Sellers, P. J., & Nickeson, J. E. (1992). Satellite remote sensing of surface energy balance: Success, failures, and unresolved issues in FIFE. *Journal of Geophysical Research*, *97*(D17), 19061–19089.

He, C. (1997). Modeling hydrologic impact of withdrawing the great lakes water for agricultural irrigation. *Journal of The American Water Resources Association*, *33*(5), 1055–1068.

He, C. (2003). Integration of GIS and simulation model for watershed management. *Environmental Modeling and Software*, *18*(8–9), 809–813.

He, C., Riggs, J. F., & Kang, Y. T. (1993). Integration of geographic information systems and a computer model to evaluate impacts of agricultural runoff on water quality. *Water Resources Bulletin*, *29*, 891–900.

He, C., Shi, C., Yang, C., & Agosti, B. P. (2001). A Windows-based GIS-AGNPS interface. *Journal of American Water Resources Association*, *37*(2), 395–406.

Hellweger, F. L., & Maidment, D. R. (1999). Definition and connection of hydrologic elements using geographic Data. *Journal of Hydro-Engineering*, *4*(1), 1–18.

Hobbins, M. T., Ramirez, J. A., Brown, T. C., & Claessens, L. J. J. M. (2001). The complementary relationship in estimation of regional evapotranspiration: The complementary relationship areal evapotranspiration and advection-aridity models. *Water Resources Research*, *37*(5), 1367–1387.

Holmes, K. W., Chadwick, O. A., & Kyriakidis, P. C. (2000). Error in a USGS 30-meter digital elevation model and its impact on terrain modeling. *Journal of Hydrology*, *233*, 154–173.

Hornberger, G. M., & Boyer, E. W. (1995). Recent advances in watershed modeling. *Review of Geophysics*, Supplement, 949–957.

Jakeman, A. J., & Hornberger, G. M. (1993). How much complexity is warranted in a rainfall-runoff model. *Water Resources Research*, *29*(8), 2637–2649.

Karvonen, T., Koivusalo, H., Jauhiainen, M., Palko, J., & Weppling, K. (1999). A hydrologic model for predicting runoff from different land use areas. *Journal of Hydrology*, *217*, 253–265.

Kenward, T., Lettenmaier, D. P., Wood, E. F., & Fielding, E. (2000). Effects of digital elevation model accuracy on hydrologic predictions. *Remote Sensing of Environment*, *74*(2000), 432–444.

Kokkonen, T. S., & Jakeman, A. J. (2001). A comparison of metric and conceptual approaches in rainfall-runoff modeling and its implications. *Water Resources Research*, *37*(9), 2345–2352.

Koren, V. I., Finnerty, B. D., Schaake, J. C., Smith, M. B., Seo, D.-J., & Duan, Q.-Y. (1999). Scale dependencies of hydrologic models to spatial variability of precipitation. *Journal of Hydrology*, *217*, 285–302.

Leavesley, G. H., & Stannard, L. G. (1995). The precipitation-runoff modeling system—PRMS. In V. Singh (Ed.), *Computer models of watershed hydrology* (pp. 281–310). Littleton, Colorado: Water Resource Publications.

Liang, X., Lettenmaier, D. P., Wood, E. F., & Burgs, S. J. (1994). A simple hydrologicly based model of land surface water and energy fluxes for generation circulation models. *Journal of Geophysical Research*, *99*, 14415–14428.

Lindsey, S. D., & Farnsworth, R. K. (1997). Sources of solar radiation estimates and their effect on daily potential evaporation for use in streamflow modeling. *Journal of Hydrology*, *201*, 348–366.

Loagues, K., & Kyriakisdis, P. C. (1997). Spatial and temporal variability in the R-5 infiltration data set: Déjé vu and rainfall-runoff simulations. *Water Resources Research*, *33*(12), 2883–2895.

Martinez, J. E., Dunchon, C. E., & Crosson, W. L. (2001). Effect of the number of soil layers on a modeled surface water budget. *Water Resources Research*, *37*(2), 367–377.

Merritt, W. S., Letcher, R. A., & Jakeman, A. J. (2003). A review of erosion and sediment transport models. *Environmental Modelling & Software*, *18*(8–9), 761–799.

Morton, F. I. (1994). Evaporation research B a critical review and its lessons for the environmental sciences. *Critical Reviews in Environment Science and Technology*, *24*(3), 237–280.

Mroczkowski, M., Raper, G. P., & Kuczera, G. (1997). The quest for more powerful validation of conceptual catchment models. *Water Resources Research*, *33*(10), 2325–2335.

Muleta, M. K., & Nicklow, J. W. (2005). Sensitivity and uncertainty analysis coupled with automatic calibration for a distributed watershed model. *Journal of Hydrology*, *306*, 127–145.

Quinn, P., Beven, K., & Culf, A. (1995). The introduction of macroscale hydrologic complexity into land surface-atmosphere transfer models and the effect on planetary boundary layer development. *Journal of Hydrology*, *166*, 421–444.

Rango, A. (1994). Application of remote sensing methods to hydrology and water resources. *Hydrological Sciences Journal*, *39*(4), 309–320.

Rango, A., & Shalaby, A. I. (1998). Operational applications of remote sensing in hydrology: Success, prospects and problems. *Hydrologic Sciences Journal*, *43*(6), 947–968.

Silberstein, R. P., Sivapalan, M., & Wyllie, A. (1999). On the validation of a coupled water and energy balance model at small catchment scales. *Journal of Hydrology*, *220*, 149–168.

Sugita, M., Usui, J., Tamagawa, I., & Kaihotsu, I. (2001). Complementary relationship with a convective boundary layer model to estimate regional evaporation. *Water Resources Research*, *37*(2), 353–365.

US Department of Agriculture. (1994). *State soil geographic (STATSGO) data base: Data use information*. Miscellaneous Publication Number 1492. Fort Worth, Texas.

US Environmental Protection Agency. (2001). *Better assessment science integrating point and nonpoint sources (BASINS) version 3.0 User's Manual*. Office of Water EPA-823-B-01-001, Washington, DC.

Vachaud, G., & Chen, T. (2002). Sensitivity of a large-scale hydrologic model to quality of input data obtained at different scales; distributed versus stochastic non-distributed modeling. *Journal of Hydrology*, *264*(2002), 101–112.

Valeo, C., & Moin, S. M. A. (2001). Hortonian and variable source area modeling in urbanizing basins. *Journal of Hydrologic Engineering*, *6*(4), 326–335.

Xu, C.-Y., & Singh, V. P. (1998). Dependence of evaporation on meteorological variables at different time-scales and intercomparison of estimation methods. *Hydrologic Processes*, *12*, 429–442.

Yan, J., & Haan, C. T. (1991). Multiobjective parameter estimation for hydrologic models-weighting of errors. *Transactions of the ASAE*, *34*(1), 135–141.

Volume 15 Issue 8 August 2007 ISSN 0967-0661

ELSEVIER

CONTROL ENGINEERING PRACTICE

A Journal of IFAC the International Federation of Automatic Control

CONTENTS

Continued on back cover

Available online at
ScienceDirect
www.sciencedirect.com

IFAC